

# Markov Switching Smooth Transition GARCH Model

N. Alemohammad<sup>\*,†</sup>, S. Rezakhah<sup>\*,†</sup> and S. Hosseinalizadeh<sup>\*, ‡</sup>

March 16, 2016

## Abstract

A Markov switching asymmetric GARCH model which imposes more leverage effect of the negative shocks is considered. The asymptotic behavior of the second moment is investigated and an upper bound for it is calculated. A bayesian strategy through Gibbs and griddy Gibbs sampling is used to estimate the parameters. Finally we study the performance of the model by two real data sets. We show that this model has the best in-sample fit via DIC and provides a better forecast when the negative skewness is large enough.

*Keywords:* Markov switching, Leverage effect, Smooth transition, DIC, Bayesian inference, Griddy Gibbs sampling.

*Mathematics Subject Classification:* 60J10, 62M10, 62F15.

## 1 Introduction

In the past four decades, dynamic financial time series based on nonlinear models has been a topic of interest. For financial time series, the ARCH and GARCH model, introduced by Engle [12] and Bollerslev [9], are surely the most popular classes of volatility models. Hamilton and Susmel [21] introduced Markov-Switching GARCH (MS-GARCH) by merging GARCH model with a hidden Markov chain, where each state of the chain allows a different GARCH behavior. Such structure exploits advantages of both, the conditional heteroscedasticity structure of GARCH models and the time-varying specifications of hidden Markov chains which improves forecasting

---

<sup>\*</sup>Department of Mathematics and Computer Science, Shahed University, Tehran, 18151/159, Iran. Email: n.alemohammad@shahed.ac.ir.

<sup>†</sup>Faculty of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran. Email: rezakhah@aut.ac.ir.

<sup>‡</sup>Department of Computer Engineering and Information Technology, Qazvin Branch, Islamic Azad University, Qazvin, Iran. Email: sasan.h.alizadeh@qiau.ac.ir.

of volatilities. Gray [17], Klaassen [23] and Haas, et al. [19] proposed some different variants of MS-GARCH models. For further studies on MS-GARCH models see Abramson and Cohen [1], Ardia [4], Alemohammad et. al [2] and Bauwens et al. [7].

One restriction of the GARCH model is its symmetry to the volatility of the past shocks. In other words, the dynamic of conditional variance in GARCH models change only with the size of square observations. This is improved by letting the conditional variance to be a function of size and sign of the lagged observations. Study of such asymmetric effects started by the work of Black [8], which studied the asymmetry of the volatility to the positive and negative shocks. This consideration is important in financial markets as there exist more volatile in response to bad news (negative shocks) than to good news (positive shocks) [16]. The asymmetric GARCH model (AGARCH) which introduced by Engle [13] captures the asymmetric effects that negative shocks have more volatile than the positive ones. Some of the main asymmetric GARCH structures which studied later on are the Exponential GARCH (EGARCH) model by Nelson [27], GJR-GARCH model by Glosten, et al.[15] and Threshold GARCH (TGARCH) model by Zakoian [29] which presented in econometric literatures. The other asymmetric structures are smooth transition models introduced by Gonzalez-Rivera [16], Medeiros and Veiga [26] and Haas et al. [20].

In this paper, we propose a generalization of the Markov switching GARCH model where the volatility of each regime is coupled with the smooth transition structure. The new model obviates the absence of asymmetric property in the Markov switching GARCH model and also the loss of transition of different levels of volatility in the smooth transition GARCH models. By considering different smooth transition models for states, the model is talent to provide a more efficient structure for modeling series with wide levels of volatilities and also encompasses the different leverage of effects causes by positive and negative shocks in each regime. As such model employ all past observations, forecasting structure could increase the complexity of the model. So we reduce the volume of calculations by proposing a dynamic programming algorithm.

We also derive necessary and sufficient conditions for stability and obtain an upper bound for the limit of the second moment by using the method of Abramson and Cohen [1] and Medeiros [26]. For the estimation of the parameters, the Bayesian inference using Gibbs and griddy Gibbs sampling is considered.

The organization of this paper is as follows: The Markov switching smooth transition GARCH model is presented in section 2. Section 3 is devoted to the statistical properties of the model. Estimation of the parameters of the model are studied in section 4. Section 5 is dedicated to the analysis of the efficiency of the proposed model by applying the model to the *S&P500* and Swiss Market index for

some special periods. Section 6 concludes.

## 2 Markov switching smooth transition GARCH model

The Markov switching smooth transition GARCH model, MS-STGARCH, for time series  $\{y_t\}$  is defined as

$$y_t = \varepsilon_t \sqrt{H_{Z_t,t}}, \quad (2.1)$$

where  $\{\varepsilon_t\}$  are iid standard normal variables,  $\{Z_t\}$  is an irreducible and aperiodic Markov chain on finite state space  $E = \{1, 2, \dots, K\}$  with transition probability matrix  $P = \|p_{ij}\|_{K \times K}$ , where  $p_{ij} = p(Z_t = j | Z_{t-1} = i)$ ,  $i, j \in \{1, \dots, K\}$ , and stationary probability measure  $\pi = (\pi_1, \dots, \pi_K)'$ . Also given that  $Z_t = j$ ,  $H_{t,j}$  (the conditional variance in regime  $j$ ) is driven by

$$H_{j,t} = a_{0j} + a_{1j}y_{t-1}^2(1 - w_{j,t-1}) + a_{2j}y_{t-1}^2w_{j,t-1} + b_jH_{j,t-1}, \quad (2.2)$$

and each of the weights  $(w_{t,j})$  is a logistic function of the past observation as

$$w_{j,t-1} = \frac{1}{1 + \exp(-\gamma_j y_{t-1})} \quad \gamma_j > 0, \quad j = 1, \dots, K, \quad (2.3)$$

which are bounded,  $0 < w_{j,t-1} < 1$ , and monotonically increasing. The parameter  $\gamma_j$  is called the slope parameter. Since  $\gamma_j > 0$ , the weight function  $0 < w_{j,t-1} < 1$  goes to zero when  $y_{t-1} \rightarrow -\infty$  and to one when  $y_{t-1} \rightarrow +\infty$ . So  $a_{1j}$  in each state will characterize negative shocks and  $a_{2j}$  positive ones. It refers to the fact that negative and positive shocks have different effects on volatility. When  $\gamma_j \rightarrow \infty$ , the logistic function becomes a step function. For  $\gamma_j$  tending to zero,  $w_{j,t-1}$  goes to 1/2 and the MS-STGARCH model tends to Markov switching GARCH model (MS-GARCH). In the case of single regime, our model is the smooth transition GARCH (STGARCH) model that is introduced by Lubrano [25].

It is assumed that  $\{\varepsilon_t\}$  and  $\{Z_t\}$  are independent. Sufficient conditions to guarantee strictly positive conditional variance are  $a_{0j}$  to be positive and  $a_{1j}, a_{2j}, b_j$  being nonnegative.

Let  $\mathcal{I}_t$  be the observation set up to time  $t$ . The conditional density function of  $y_t$  given past information is obtained as follows:

$$f(y_t | \mathcal{I}_{t-1}) = \sum_{i=1}^K \alpha_i^{(t)} \phi\left(\frac{y_t}{\sqrt{H_{i,t}}}\right) \quad (2.4)$$

in which  $\alpha_i^{(t)} = p(Z_t = i | \mathcal{I}_{t-1})$  (that is obtained in the next section), and  $\phi(\cdot)$  is the probability density function of the standard normal distribution.

### 3 Statistical Properties of the model

In this section, the statistical properties of the MS-STGARCH model are investigated and the conditional variance of the process is obtained. We show that the model, under some conditions on coefficients and transition probabilities, is asymptotically stable in the second moment. An appropriate upper bound for the limiting value of the second moment is obtained.

#### 3.1 Forecasting

The conditional variance of MS-STGARCH model is given by

$$\begin{aligned} Var(Y_t | \mathcal{I}_{t-1}) &= \sum_{i=1}^K \alpha_i^{(t)} H_{i,t} = \sum_{i=1}^K \alpha_i^{(t)} (a_{0i} + a_{1i} y_{t-1}^2 (1 - w_{i,t-1}) \\ &\quad + a_{2i} y_{t-1}^2 w_{i,t-1} + b_i H_{i,t-1}) \end{aligned} \quad (3.5)$$

as  $H_{t,i}$  is the conditional variance of  $i$ -th state. This relation shows that the conditional variance of this model is affected by changes in regime and conditional variance of each state.

As using all past observations for forecasting could increase the complexity of the model, we reduce the volume of calculations by proposing a dynamic programming algorithm. At each time  $t$ ,  $\alpha_i^{(t)}$  (in equation (2.4), (3.5)) can be obtained from a dynamic programming method based on the forward recursion algorithm, proposed in remark (3.1).

**Remark 3.1** *The value of  $\alpha_j^{(t)}$  is obtained recursively by*

$$\alpha_j^{(t)} = \frac{\sum_{m=1}^K f(y_{t-1} | Z_{t-1} = m, \mathcal{I}_{t-2}) p(Z_{t-1} = m | \mathcal{I}_{t-2}) p_{m,j}}{\sum_{m=1}^K f(y_{t-1} | Z_{t-1} = m, \mathcal{I}_{t-2}) p(Z_{t-1} = m | \mathcal{I}_{t-2})}. \quad (3.6)$$

**Proof 3.1** *See Appendix A.*

### 3.2 Stability

In this subsection, we investigate the stability of second moment of the MS-STGARCH model. Similar to our previous work [2], we are looking for an upper bound of our model for the second moment. Assume  $M$  is a positive constant and let

$$\Omega = [a_{01} + |a_{21} - a_{11}|M^2, \dots, a_{0K} + |a_{2K} - a_{1K}|M^2]', \quad (3.7)$$

be a vector with  $K$  component,  $C$  denotes a  $K^2$ -by- $K^2$  block matrix as

$$C = \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{K1} \\ C_{12} & C_{22} & \cdots & C_{K2} \\ \vdots & & & \vdots \\ C_{1K} & C_{2K} & \cdots & C_{KK} \end{pmatrix} \quad (3.8)$$

with each block given by

$$C_{jk} = p(Z_{t-1} = j | Z_t = k)(ue'_j + v), \quad j, k = 1, \dots, K, \quad (3.9)$$

where  $u = [a_{11} + (\delta + \frac{1}{2})|a_{21} - a_{11}|, \dots, a_{1K} + (\delta + \frac{1}{2})|a_{2K} - a_{1K}|]'$ ,  $e_j$  is a  $K$ -by-1 vector of all zeros, except its  $j$ th element, which is one, and  $v$  is a diagonal  $K$ -by- $K$  matrix with elements  $[b_1, \dots, b_K]$  on its diagonal.

Let  $\Pi = [\pi_1 e'_1, \dots, \pi_K e'_K]$  and consider  $\rho(A)$  denotes the spectral radius of a matrix  $A$ , then we have the following theorem for the stationarity condition of the MS-STGARCH model.

**Theorem 3.1** *Let  $\{Y_t\}_{t=0}^\infty$  follows the MS-STGARCH model, defined by (2.1)-(2.3), the process is asymptotically stable in variance and  $\lim_{t \rightarrow \infty} E(Y_t^2) \leq \Pi'(I - C)^{-1}\Omega$ , if  $\rho(C) < 1$ .*

**Proof 3.2** *See Appendix B.*

## 4 Estimation

For the estimation of parameters, we apply the Bayesian inference, that is extensively used in literature ([6], [7] and [25]).

Let  $Y_t$  be the vector  $(y_1, \dots, y_t)$ . In what follows, for the case of two regimes, we are going to estimate the  $Z_t = (z_1, \dots, z_t)$  and the parameter vectors  $\eta = (\eta_{11}, \eta_{22})$  and  $\theta = (\theta_1, \theta_2)$ , where  $\theta_k = (a_{0k}, a_{1k}, a_{2k}, b_k, \gamma_k)$  for  $k = 1, 2$  from the posterior density

$$p(\theta, \eta, Z|Y) \propto p(\theta, \eta)p(Z|\theta, \eta)f(Y|\theta, \eta, Z), \quad (4.10)$$

in which  $Y = (y_1, \dots, y_T)$ ,  $Z = (z_1, \dots, z_T)$  and  $p(\theta, \eta)$  is the prior of the parameters. The conditional probability mass function of  $Z$  given the  $(\theta, \eta)$  is independent of  $\theta$ , so

$$\begin{aligned} p(Z|\theta, \eta) &= p(Z|\eta_{00}, \eta_{11}) \\ &= \prod_{t=1}^T p(z_{t+1}|z_t, \eta_{00}, \eta_{11}) \\ &= p_{00}^{n_{00}}(1 - p_{00})^{n_{01}}p_{11}^{n_{11}}(1 - p_{11})^{n_{10}}, \end{aligned} \quad (4.11)$$

where  $n_{ij} = \#\{z_t = j|z_{t-1} = i\}$  (the number of transitions from regime  $i$  to regime  $j$ ). The conditional density function of  $Y$  given the realization of  $Z$  and the parameters is factorized in the following way:

$$f(Y|\eta, \theta, Z) = \prod_{t=1}^T f(y_t|\theta, z_t = k, Y_{t-1}), \quad k = 1, 2, \quad (4.12)$$

where the one step ahead predictive densities are:

$$f(y_t|\theta, z_t = k, Y_{t-1}) = \frac{1}{\sqrt{2\pi H_{k,t}}} \exp\left(-\frac{y_t^2}{H_{k,t}}\right). \quad (4.13)$$

Since the straight sampling from the posterior density (4.10) is not possible, we apply the Gibbs sampling algorithm with three blocks:  $\theta$ ,  $\eta$  and  $Z$ .

A brief description of the Gibbs algorithm: Suppose the superscript  $(r)$  on a parameter is a sample of the parameter at the  $r$ -th iteration of the algorithm. At any iteration of the algorithm, three steps will be done:

- (i) Draw the random sample of the state variable  $Z^{(r)}$  given  $\eta^{(r-1)}$ ,  $\theta^{(r-1)}$ .
- (ii) Draw the random sample of the transition probabilities  $\eta^{(r)}$  given  $Z^{(r)}$ .
- (iii) Draw the random sample of the  $\theta^{(r)}$  given  $Z^{(r)}$  and  $\eta^{(r)}$ .

These steps are repeated until the convergency is obtained. In what follows the sampling of each block are explained.

#### 4.1 Sampling $z_t$

This step is devoted to the sampling of  $p(z_t|\eta, \theta, Y_t)$  that is performed by Chib[10], (see also [22]). Suppose  $p(z_1|\eta, \theta, Y_0)$  be the stationary distribution of the chain,

$$p(z_t|\eta, \theta, Y_t) \propto f(y_t|\theta, z_t = k, Y_{t-1})p(z_t|\eta, \theta, Y_{t-1}), \quad (4.14)$$

where the predictive density  $f(y_t|\theta, z_t = k, Y_{t-1})$  is calculated by the relation (4.13) and by the law of total probability  $p(z_t|\eta, \theta, Y_{t-1})$  is given by:

$$p(z_t|\eta, \theta, Y_{t-1}) = \sum_{z_{t-1}=1}^K p(z_{t-1}|\eta, \theta, Y_{t-1})\eta_{z_{t-1}z_t}. \quad (4.15)$$

Given the filter probabilities  $(p(z_t|\eta, \theta, Y_t))$ , we run a backward algorithm, starting from  $t = T$  that  $z_T$  is derived from  $p(z_T|\eta, \theta, Y)$ . For  $t = T-1, \dots, 0$  the sample is derived from  $p(z_t|z_{t+1}, \dots, z_T, \theta, \eta, Y)$ , which is obtained by

$$p(z_t|z_{t+1}, \dots, z_T, \theta, \eta, Y) \propto p(z_t|\eta, \theta, Y_t)\eta_{z_t, z_{t+1}}.$$

To derive  $z_t$  from  $p(z_t|\cdot) = p_{z_t}$  is by sampling from the conditional probabilities  $q_j = p(Z_t = j|Z_t \geq j, \cdot)$  which are given by

$$p(Z_t = j|Z_t \geq j, \cdot) = \frac{p_j}{\sum_{l=j}^K q_l}.$$

After generating a uniform (0,1) number  $U$ , if  $U \leq q_j$  then  $z_t = j$ , otherwise increase  $j$  to  $j+1$  and generate another uniform (0,1) and compare it by  $q_{j+1}$ .

## 4.2 Sampling $\eta$

This stage is devoted to sample  $\eta = (\eta_{11}, \eta_{22})$  from the posterior probability  $p(\eta|\theta, Y_t, Z_t)$  that is independent of  $Y_t, \theta$ . We consider independent beta prior density for each of  $\eta_{11}$  and  $\eta_{22}$ . For example,

$$p(\eta_{11}|Z_t) \propto p(\eta_{11})p(Z_t|\eta_{11}) = \eta_{11}^{c_{11}+n_{11}-1}(1-\eta_{11})^{c_{12}+n_{12}-1},$$

where  $c_{11}$  and  $c_{12}$  are the parameters of beta prior,  $n_{ij}$  is the number of transition from  $z_{t-1} = i$  to  $z_t = j$ . In the same way the sample of  $\eta_{22}$  is obtained.

## 4.3 Sampling $\theta$

The posterior density of  $\theta$  given the prior  $p(\theta)$  is given by:

$$p(\theta|Y, Z, \eta) \propto p(\theta) \prod_{t=1}^T f(y_t|\theta, z_t = k, Y_{t-1}) = p(\theta) \prod_{t=1}^T \frac{1}{\sqrt{2\pi H_{k,t}}} \exp\left(-\frac{y_t^2}{H_{k,t}}\right), \quad (4.16)$$

which is independent of  $\eta$ . To sample from the  $p(\theta|Y, Z, \eta)$  we use the Griddy Gibbs algorithm that introduced by Ritter and Tanner [28]. This method is very applicable

in researches (for example [5], [6] and [7]). Given samples at iteration  $r$  the Griddy Gibbs at iteration  $r + 1$  proceeds as follows:

1. Select a grid of points, such as  $a_{0i}^1, a_{0i}^2, \dots, a_{0i}^G$ . Using (4.16), evaluate the conditional posterior density function  $k(a_{0i}|Z_t, Y_t, \theta_{-a_{0i}})$  over the grid points to obtain the vector  $G_k = (k_1, \dots, k_G)$ .
2. By a deterministic integration rule using the G points, compute  $G_\Phi = (0, \Phi_2, \dots, \Phi_G)$  with

$$\Phi_j = \int_{a_{0i}^1}^{a_{0i}^j} k(a_{0i}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t) da_{0i}, \quad i = 2, \dots, G. \quad (4.17)$$

3. Simulate  $u \sim U(0, \Phi_G)$  and invert  $\Phi(a_{0i}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t)$  by numerical interpolation to obtain a sample  $a_{0i}^{(r+1)}$  from  $a_{0i}|\theta_{-a_{0i}}^{(r)}, Z_t^{(r)}, Y_t$ .
4. Repeat steps 1-3 for other parameters.

For the prior densities of all elements of  $\theta$ , it can be considered independent uniform densities over the finite intervals.

## 5 Two sets of empirical data

In this section we apply the daily log returns of the *S&P500* and Swiss Market Index (SMI). The sample period of *S&P500* is from 03/01/2005 to 03/11/2014 (2500 observations), the first 2000 observations are employed to estimate the parameters while the remaining 500 are used in a forecasting analysis. The selected interval of SMI is from 10/02/2004 to 11/11/2015 (3000 observations), the first 2500 indexes are used to estimate and the rest of them for forecasting. Figure 1 demonstrates the percentage returns of <sup>1</sup> of the *S&P500* and SMI. In Table 1, we report summary descriptive statistics of returns. We observe the means are close to zero and also a slightly negative skewness and a common excess kurtosis for each data set.

### 5.1 Estimation

Using the Bayesian inference, we estimate the parameters of STGARCH [25], two-state MS-GARCH [19] and MS-STGARCH with two regimes. The prior density of  $\eta_{11}$  and  $\eta_{22}$  (the transition probability from one regime to other regime) are drawn from the beta distribution and the others are assumed to be uniform over a finite intervals. The number of iterations of Gibbs algorithm for each model is 10000 which half of them are burn-in-phase. The posterior means and standard deviations

---

<sup>1</sup>Percentage return is defined as  $r_t = 100 * \log(\frac{P_t}{P_{t-1}})$ , where  $P_t$  is the index level at time  $t$ .



for the models corresponding to *S&P500* and *SMI* are reported in Table 2 and Table 3 show that the standard deviations are small enough in most cases. The single-regime me STGARCH has different reactions to negative and positive shocks aren't flexible like Markov switching model to transit between low volatility and high volatility and reverse. According to results we find that in the MS-GARCH and MS-STGARCH models the second regimes have higher volatility. MS-GARCH defined by setting  $\alpha_{2j}$  and  $\gamma_j$ ,  $j = 1, 2$  equal to zero in relation (2.2) that implies a symmetric response respect to positive and negative shocks of the same magnitude. The MS-STGARCH encompasses two previlages: the transition between low and high volatility states and the different reactions to positive and negative shocks. Since in the MS-STGARCH,  $a_{11} > a_{21}$  and  $a_{12} > a_{22}$  and according to structure of  $w_{j,t-1}$  (2.3) we deduce that the negative shocks have more affect on volatility than the positive shocks in two regimes. The obvious differences between  $\gamma_1$  and  $\gamma_2$  in Table 2 and 3 expresses that the leverage effect in each regime is different. In all cases the posterior means of transition probability ( $\eta_{11}$  and  $\eta_{22}$ ) indicate scarce mixing between regimes. Smoothed probabilities for the second regimes of MS-STGARCH according to *S&P500* and *SMI* log returns are plotted in Figure 2.

## 5.2 Deviance information criterion

In order to evaluate the goodness of fit of models, we apply the deviance information criterion introduced by Spiegelhalter et al. see also [14]. The smallest DIC determines the best fitting model.

In the Markov-switching models, the likelihood is calculated by the following formula:

$$f(Y|\Theta) = \prod_{t=1}^T f(y_t|\mathcal{I}_{t-1}, \Theta)$$

in which  $\Theta$  is the vector of all parameters in model and  $f(y_t|\mathcal{I}_{t-1})$  is obtained from (2.4). The deviance information criterion is computed as:

$$DIC = 2 \log(f(Y|\hat{\Theta})) - 4E_{Y|\Theta}[\log(f(Y|\Theta))], \quad (5.18)$$

that  $\hat{\Theta}$  is the posterior means of the vector  $\Theta$ . The results concerning DIC are reported in Table 4. It is apparent that the DIC of MS-STGARCH is the smallest value in both data sets.

## 5.3 Forecasting performance

For appraising the MS-STGARCH model to forecast the future treatment, we survey the one-day-ahead value at risk (VaR) forecasts. The one-day-ahead VaR at risk

level  $\alpha \in (0, 1)$ ,  $\text{VaR}(\alpha)$  is obtained by calculating the  $(1 - \alpha)$ th percentile of the one-day-ahead predictive distribution [4]. Based on the last 500 returns (of *S&P500* and *SMI*), the out of sample VaR measures for each model are calculated. To test the VaR values, we investigate the random sequence  $\{v_t(\alpha)\}$  that

$$V_t(\alpha) = \begin{cases} I\{y_{t+1} < \text{VaR}(\alpha)\} & \text{if } \alpha > 0.5 \\ I\{y_{t+1} > \text{VaR}(\alpha)\} & \text{if } \alpha \leq 0.5. \end{cases}$$

The out-of-sample VaR measures has the good performance if  $\{v_t(\alpha)\}$  is an independent and also follows the bellow distribution (unconditional coverage):

$$V_t(\alpha) \sim \begin{cases} \text{Bernoulli}(1 - \alpha) & \text{if } \alpha > 0.5 \\ \text{Bernoulli}(\alpha) & \text{if } \alpha \leq 0.5, \end{cases}$$

The statistics for the independency ( $LR_{ind}$ ) and unconditional coverage ( $LR_{uc}$ ) asymptotically are Chi-squared distributed with one degree of freedom. By adding the  $LR_{ind}$  and  $LR_{uc}$ , a dual test that inquires both hypotheses (independency and unconditional coverage), i.e. conditional coverage (see [11] and [4]):

$$LR_{cc} = LR_{ind} + LR_{uc},$$

$LR_{cc}$  is  $\chi^2$  distributed with two degrees of freedom. The lower value of  $LR_{cc}$  than the critical  $\chi^2$  distribution means that the correct conditional coverage and as a result the good VaR forecasts.

The results of the tests are reported in Table 5 for *S&P500* and Table 6 for *SMI*. The second and third columns demonstrate the theoretical expected violations and the number of empirical violations. The  $LR_{ind}$  statistics in Table 5 and 6 are smaller than critical  $\chi^2$  distribution for all cases and so independency test isn't rejected. The results of Table 5 show that the real number of violations for the MS-STGARCH are closer to expected values than for the STGARCH and MS-GARCH except for  $\alpha = 0.99$ , which can be considered as a better performance. At the 10% significance level, the  $LR_{uc}$  test is rejected three times for the STGARCH, two times for MS-GARCH and is accepted for all quantities of  $\alpha$  for the MS-STGARCH. Also the values of  $LR_{cc}$  are lower than the critical  $\chi^2$  distribution for MS-STGARCH, while this test are rejected three times for the STGARCH and one time for the MS-GARCH. In Table 6 the  $LR_{uc}$  is larger than critical value three times for STGARCH and two times for the MS-GARCH and MS-STGARCH. Also the conditional coverage test is rejected thrice for the STGARCH, once for the MS-GARCH and twice for the MS-STGARCH. The results indicate that the new model has the better forecasting performance than the STGARCH model in both examples. In the case of greater negative skewness (*S&P500* returns), the out-of-sample VaR forecasts of the MS-STGARCH has better performance than the MS-GARCH.

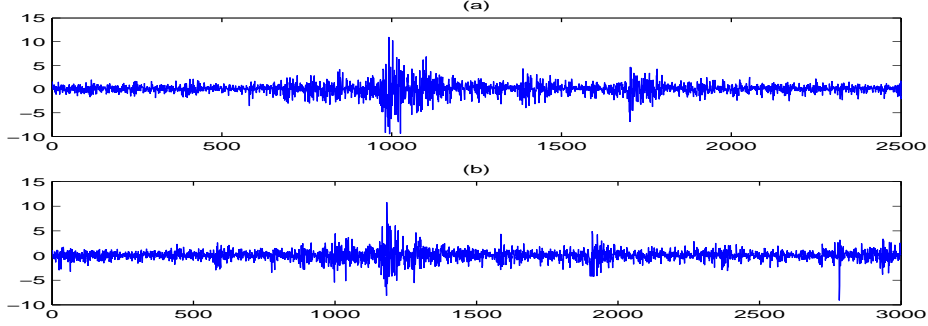


Figure 1: Percentage daily log returns of (a): *S&P500* data, (b): Swiss Market Index (SMI).

Table 1: Descriptive statistics for the *S&P500* and Swiss market index daily log returns.

	Mean	Std. dev.	Skewness	Maximum	Minimum	Kurtosis
SandP500	0.023	1.287	-0.337	10.957	-9.469	14.049
Swiss Market	0.015	1.222	-0.263	10.723	-9.085	12.162

Table 2: Posterior means and standard deviations (*S&P500* daily log returns).

	MS-STGARCH		MS-GARCH		ST-GARCH	
	Mean	Std.dev.	Mean	Std.dev	Mean	Std.dev
$a_{01}$	0.194	0.001	0.233	0.004	0.336	0.011
$a_{11}$	.276	0.008	0.278	0.012	0.421	0.019
$a_{21}$	0.085	0.006	0	0	0.121	0.016
$b_1$	0.289	0.003	0.320	0.009	0.364	0.014
$\gamma_1$	2.345	0.132	0	0	2.206	0.218
$a_{02}$	0.717	0.087	0.779	0.012	-	-
$a_{12}$	0.677	0.007	0.430	0.011	-	-
$a_{22}$	0.365	0.013	0	0	-	-
$b_2$	0.264	0.015	0.207	0.007	-	-
$\gamma_2$	1.097	0.017	0	0	-	-
$\eta_{11}$	0.986	0.004	0.994	0.003	-	-
$\eta_{22}$	0.985	0.005	0.991	0.004	-	-

Table 3: Posterior means and standard deviations (Swiss market index daily log returns).

	MS-STGARCH		MS-GARCH		ST-GARCH	
	Mean	Std.dev.	Mean	Std.dev	Mean	Std.dev
$a_{01}$	0.296	0.012	0.287	0.008	0.432	0.016
$a_{11}$	0.450	0.021	0.292	0.011	0.465	0.019
$a_{21}$	0.155	0.017	0	0	0.177	0.013
$b_1$	0.406	0.013	0.393	0.011	0.325	0.011
$\gamma_1$	1.397	0.136	0	0	2.719	0.282
$a_{02}$	0.876	0.018	0.819	0.019	-	-
$a_{12}$	0.516	0.023	0.385	0.016	-	-
$a_{22}$	0.257	0.014	0	0	-	-
$b_2$	0.155	0.008	0.127	0.009	-	-
$\gamma_2$	2.107	0.196	0	0	-	-
$\eta_{11}$	0.995	0.002	0.994	0.003	-	-
$\eta_{22}$	0.981	0.009	0.981	0.008	-	-

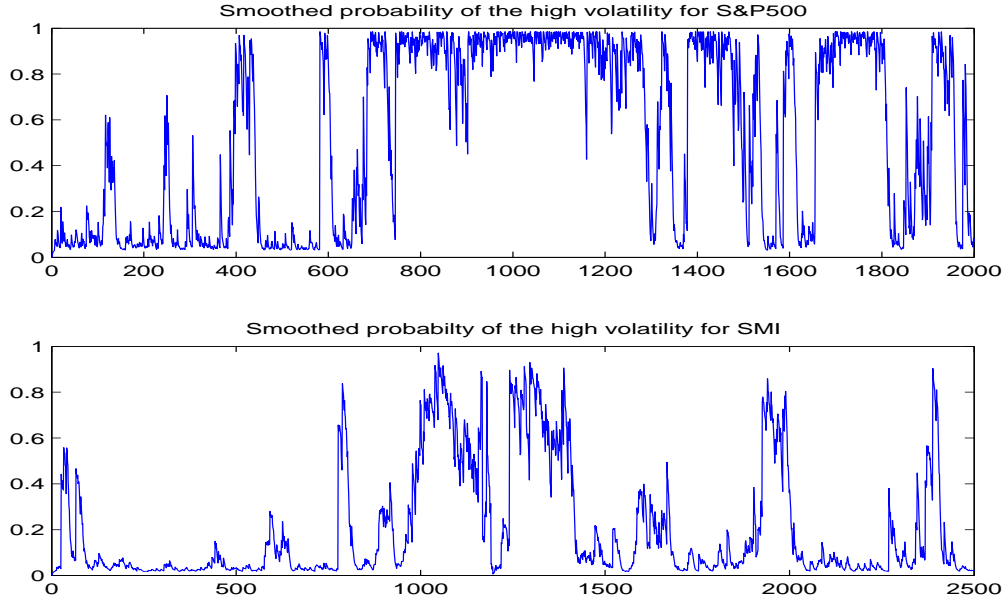


Figure 2: Smoothed probabilities of the second regimes for the MS-STGARCH in S&P500 and SMI.

Table 4: Deviance information criterion (DIC)

Model	S&P500 returns	Swiss market returns
ST-GARCH	7513.5	8339.2
MS-GARCH	7257.1	8275.3
MS-STGARCH	7147.8*	8210.6*

Table 5: VaR results of *S&P500* daily log returns.

Model	$\alpha$	$E(V_t(\alpha))$	N	UC	IND	CC
STGARCH	0.99	5	5	0	0.101	0.101
	0.95	25	20	1.147	1.582	2.729
	0.9	50	31	9.235	0.5317	9.767
	0.1	50	28	12.684	1.191	13.8757
	0.05	25	8	16.441	0.260	16.701
	0.01	5	2	2.365	0.102	2.381
MS-GARCH	0.99	5	8	1.526	0.260	1.786
	0.95	25	27	0.156	0.1787	0.335
	0.9	50	42	1.531	0.827	2.358
	0.1	50	37	4.112	0.252	4.364
	0.05	25	15	4.926	0.539	5.465
	0.01	5	2	2.365	0.0161	2.3811
MS-STGARCH	0.99	5	9	2.596	0.408	3.004
	0.95	25	27	0.156	0.261	0.417
	0.9	50	45	0.595	0.301	0.897
	0.1	50	44	0.9	0.252	1.109
	0.05	25	18	2.3	0.2746	2.575
	0.01	5	2	2.38	0.016	2.36

Table 6: VaR results of SMI daily log returns.

Model	$\alpha$	$E(V_t(\alpha))$	N	UC	IND	CC
STGARCH	0.99	5	7	0.718	0.199	0.917
	0.95	25	19	1.646	1.505	3.15
	0.9	50	36	4.779	0.171	4.95
	0.1	50	32	8.148	0.7421	8.89
	0.05	25	14	6.017	0.808	6.820
	0.01	5	5	0	0.101	0.101
MS-GARCH	0.99	5	8	1.538	0.261	1.794
	0.95	25	23	0.173	0.0.004	0.177
	0.9	50	43	1.138	0.497	1.635
	0.1	50	39	2.888	2.106	4.99
	0.05	25	17	3.021	1.199	4.22
	0.01	5	5	0	0.101	.101
MS-STGARCH	0.99	5	8	1.538	0.261	1.799
	0.95	25	21	0.711	1.846	2.557
	0.9	50	44	0.830	0.362	1.193
	0.1	50	35	5.52	1.247	6.77
	0.05	25	16	6.018	0.808	6.82
	0.01	5	5	0	0.101	0.101

## 6 Conclusion

The MS-STGARCH model extend the MS-GARCH model by considering the different smooth transition structures in each regime. It is also an extension of the STGARCH model that transits between the different levels of volatilities. The proposed model authorizes for greater flexibility to capture volatility clustering and leverage effect that are usual patterns in financial time series. For the estimation of parameters we apply the Bayesian estimation algorithm. We present a simple necessary and sufficient condition for the existence of an upper bound for the second moment. For the estimation of the parameters the Bayesian inference is used by applying the Griddy Gibbs algorithm. We fit STGARCH, MS-GARCH and MS-STGARCH models to the *S&P500* and SMI log returns. We demonstrate that when the negative skewness coefficient is large enough our model has the best VaR forecasts among other models.

This model has the potential to be applied for modeling and forecasting conditional volatility of financial time series . Further researches could be oriented to investigate the existence of the third and the fourth moments of the process and derive the necessary and sufficient conditions for stationarity and ergodicity properties of the process. For the sake of simplicity it was assumed that the process conditional mean is zero, this assumption could be relaxed by refining the structure of model to allow ARMA structure for conditional mean. Finally it might be interesting to replace the Gaussian distribution with t-distribution and investigate the ability for modeling heavy tailed property of financial time series.

## Appendix A

Proof of Remark 3.1.

As the hidden variables  $\{Z_t\}_{t \geq 1}$  have Markov structure in MS-STARCH model, so

$$\begin{aligned}
\alpha_j^{(t)} &= p(Z_t = j | \mathcal{I}_{t-1}) = \sum_{m=1}^K P(Z_t = j, Z_{t-1} = m | \mathcal{I}_{t-1}) \\
&= \sum_{m=1}^K p(Z_t = j | Z_{t-1} = m, \mathcal{I}_{t-1}) p(Z_{t-1} = m | \mathcal{I}_{t-1}) \\
&= \sum_{m=1}^K p(Z_t = j | Z_{t-1} = m) p(Z_{t-1} = m | \mathcal{I}_{t-1}) \\
&= \frac{\sum_{m=1}^K f(\mathcal{I}_{t-1}, Z_{t-1} = m) p_{m,j}}{\sum_{m=1}^K f(\mathcal{I}_{t-1}, Z_{t-1} = m)} \\
&= \frac{\sum_{m=1}^K f(y_{t-1} | Z_{t-1} = m, \mathcal{I}_{t-2}) p(Z_{t-1} = m | \mathcal{I}_{t-2}) p_{m,j}}{\sum_{m=1}^K f(y_{t-1} | Z_{t-1} = m, \mathcal{I}_{t-2}) p(Z_{t-1} = m | \mathcal{I}_{t-2})}, \tag{6.19}
\end{aligned}$$

where

$$f(y_{t-1} | Z_{t-1} = m, \mathcal{I}_{t-2}) = \phi\left(\frac{y_{t-1}}{\sqrt{H_{m,t-1}}}\right).$$

## Appendix B

Proof of Theorem 3.1.

The second moment of the model can be calculated as:

$$\begin{aligned}
E(y_t^2) &= E(H_{Z_t,t}) = E_{Z_t}[E_{t-1}(H_{Z_t,t} | z_t)] \\
&= \sum_{z_t=1}^K \pi_{z_t} E_{t-1}(H_{Z_t,t} | z_t). \tag{6.20}
\end{aligned}$$

$E_t(\cdot)$  denotes the expectation with respect to the information up to time  $t$ . Also for summarization, we shall use  $E(\cdot | z_t)$  and  $p(\cdot | z_t)$  to represent  $E(\cdot | Z_t = z_t)$  and  $P(\cdot | Z_t = z_t)$ , respectively, where  $z_t$  is the realization of the state at time  $t$ . The

conditional variance under the chain state,  $m$ , is investigated as follows:

$$\begin{aligned}
E_{t-1}(H_{m,t}|z_t) &= E_{t-1}(a_{0m} + a_{1m}y_{t-1}^2(1 - w_{m,t-1}) + a_{2m}y_{t-1}^2w_{m,t-1} + b_m H_{m,t-1}|z_t) \\
&= \underbrace{a_{0m}}_I + \underbrace{a_{1m}E_{t-1}[y_{t-1}^2|z_t]}_{II} + \underbrace{(a_{2m} - a_{1m})E_{t-1}[y_{t-1}^2w_{m,t-1}|z_t]}_{III} \\
&\quad + \underbrace{b_mE_{t-1}[H_{m,t-1}|z_t]}_{IV}. \tag{6.21}
\end{aligned}$$

The term (II) in (6.21) can be interpreted as follows:

$$\begin{aligned}
E_{t-1}[y_{t-1}^2|z_t] &= \sum_{z_{t-1}=1}^K \int_{S_{\mathcal{I}_{t-1}}} y_{t-1}^2 p(\mathcal{I}_{t-1}|z_t, z_{t-1}) p(z_{t-1}|z_t) d\mathcal{I}_{t-1} \\
&= \sum_{z_{t-1}=1}^K p(z_{t-1}|z_t) E_{t-2}[H_{Z_{t-1}, t-1}|z_{t-1}], \tag{6.22}
\end{aligned}$$

where  $S_{\mathcal{I}_{t-1}}$  is the support of  $\mathcal{I}_{t-1} = (y_1, \dots, y_{t-1})$ .

**Upper bound for III in (6.21):** Let  $0 < M < \infty$  be a constant, so

$$\begin{aligned}
E_{t-1}[y_{t-1}^2w_{m,t-1}|z_t] &= E_{t-1}[y_{t-1}^2w_{m,t-1}I_{|y_{t-1}| < M}|z_t] \\
&\quad + E_{t-1}[y_{t-1}^2w_{m,t-1}I_{|y_{t-1}| \geq M}|z_t]
\end{aligned}$$

in which

$$I_{x < a} = \begin{cases} 1 & \text{if } x < a \\ 0 & \text{otherwise.} \end{cases}$$

As by (2.3),  $0 < w_{m,t-1} < 1$  and so

$$E_{t-1}[y_{t-1}^2w_{m,t-1}|z_t] \leq M^2 + E_{t-1}[y_{t-1}^2w_{m,t-1}I_{|y_{t-1}| \geq M}|z_t],$$

also

$$\begin{aligned}
E_{t-1}[y_{t-1}^2w_{m,t-1}I_{|y_{t-1}| \geq M}|z_t] &= \int_{S_{\mathcal{I}_{t-2}}, y_{t-1} \leq -M} y_{t-1}^2[w_{m,t-1}] p(\mathcal{I}_{t-1}|z_t) d\mathcal{I}_{t-1} \\
&\quad + \int_{S_{\mathcal{I}_{t-2}}, y_{t-1} \geq M} y_{t-1}^2[w_{m,t-1}] p(\mathcal{I}_{t-1}|z_t) d\mathcal{I}_{t-1},
\end{aligned}$$

by (2.3),

$$\lim_{y_{t-1} \rightarrow +\infty} w_{m,t-1} = 1 \quad \lim_{y_{t-1} \rightarrow -\infty} w_{m,t-1} = 0, \tag{6.23}$$

therefore according to the definition of limit at infinity, for a small number  $\delta > 0$ , there will exist a finite constant  $M > 0$  such that if  $y_{t-1} \geq M$ ,  $|w_{m,t-1} - 1| \leq \delta$  and if  $y_{t-1} \leq -M$ ,  $|w_{m,t-1}| \leq \delta$ . Hence

$$\begin{aligned} E_{t-1}[y_{t-1}^2 w_{m,t-1} I_{|y_{t-1}| \geq M} | z_t] &\leq \delta \int_{S_{\mathcal{I}_{t-2}}, y_{t-1} \leq -M} y_{t-1}^2 p(\mathcal{I}_{t-1} | z_t) d\mathcal{I}_{t-1} \\ &\quad + (\delta + 1) \int_{S_{\mathcal{I}_{t-2}}, y_{t-1} \geq M} y_{t-1}^2 p(\mathcal{I}_{t-1} | z_t) d\mathcal{I}_{t-1}. \end{aligned}$$

Since the distribution of the  $\{\varepsilon_t\}$  is symmetric, then

$$\begin{aligned} \delta \int_{S_{\mathcal{I}_{t-2}}, y_{t-1} \leq -M} y_{t-1}^2 p(\mathcal{I}_{t-1} | z_t) d\mathcal{I}_{t-1} &\leq \delta \int_{S_{\mathcal{I}_{t-2}}, -\infty < y_{t-1} < 0} y_{t-1}^2 p(\mathcal{I}_{t-1} | z_t) d\mathcal{I}_{t-1} \\ &= \delta \frac{E_{t-1}[y_{t-1}^2 | z_t]}{2} \end{aligned}$$

and

$$\begin{aligned} (\delta + 1) \int_{S_{\mathcal{I}_{t-2}}, y_{t-1} \geq M} y_{t-1}^2 p(\mathcal{I}_{t-1} | z_t) d\mathcal{I}_{t-1} &\leq (\delta + 1) \int_{S_{\mathcal{I}_{t-2}}, 0 < y_{t-1} < \infty} y_{t-1}^2 p(\mathcal{I}_{t-1} | z_t) d\mathcal{I}_{t-1} \\ &= (\delta + 1) \frac{E_{t-1}[y_{t-1}^2 | z_t]}{2}. \end{aligned}$$

Therefor

$$E_{t-1}[y_{t-1}^2 w_{m,t-1} | z_t] \leq M^2 + (\delta + \frac{1}{2}) E_{t-1}[y_{t-1}^2 | z_t].$$

**Upper bound for IV in (6.21):**

$$\begin{aligned} b_m E_{t-1}(H_{m,t-1} | z_t) &= b_m \int_{S_{\mathcal{I}_{t-1}}} H_{m,t-1} p(\mathcal{I}_{t-1} | z_t) d\mathcal{I}_{t-1} \\ &= b_m \sum_{z_{t-1}=1}^K p(z_{t-1} | z_t) E_{t-2}(H_{m,t-1} | z_{t-1}). \end{aligned} \tag{6.24}$$

By replacing the obtained upper bounds and relations (6.22) in (6.21), the upper



bound for  $E_{t-1}(H_{m,t}|z_t)$  is obtained by:

$$\begin{aligned}
E_{t-1}(H_{m,t}|z_t) &\leq a_{0m} + |a_{2m} - a_{1m}|M^2 \\
&+ \sum_{z_{t-1}=1}^K [a_{1m} + |a_{2m} - a_{1m}|(\delta + \frac{1}{2})]p(z_{t-1}|z_t)E_{t-2}[H_{z_{t-1},t-1}|z_{t-1}] \\
&+ \sum_{z_{t-1}=1}^K b_m p(z_{t-1}|z_t)E_{t-2}(H_{m,t-1}|z_{t-1}),
\end{aligned} \tag{6.25}$$

in which by Bayes' rule

$$p(z_{t-i}|z_t) = \frac{\pi_{z_{t-i}}}{\pi_{z_t}} \{P_{z_{t-i}z_t}\},$$

where  $P$  is the transition probability matrix.

Let  $A_t(j, k) = E_{t-1}[H_{j,t}|Z_t = k]$ ,  $A_t = [A_t(1, 1), A_t(2, 1), \dots, A_t(K, 1), A_t(1, 2), \dots, A_t(K, K)]$  be a  $K^2$ -by-1 vector and consider  $\dot{\Omega} = (\Omega', \dots, \Omega')'$  be a vector that is made of  $K$  vector  $\Omega$ . By (3.7)-(3.9), the following recursive inequality is attained,

$$\mathbf{A}_t \leq \dot{\Omega} + \mathbf{C}\mathbf{A}_{t-1}, \quad t \geq 0. \tag{6.26}$$

with some initial conditions  $\mathbf{A}_{-1}$ . The relation (6.26) implies that

$$A_t \leq \dot{\Omega} \sum_{i=0}^{t-1} C^i + C^t A_0 := B_t. \tag{6.27}$$

Following the matrix convergence theorem [24], the necessary condition for the convergence of  $B_t$  when  $t \rightarrow \infty$  is that  $\rho(C) < 1$ . Under this condition,  $C^t$  converges to zero as  $t$  goes to infinity and  $\sum_{i=0}^{t-1} C^i$  converges to  $(I - C)^{-1}$  provided that matrix  $(I - C)$  is invertible. So if  $\rho(C) < 1$ ,

$$\lim_{t \rightarrow \infty} A_t \leq (I - C)^{-1} \dot{\Omega}.$$

By (6.20) the upper bound for the asymptotic behavior of unconditional variance is given by

$$\lim_{t \rightarrow \infty} E(y_t^2) \leq \Pi'(I - C)^{-1} \dot{\Omega}.$$

## References

- [1] Abramson, A. Cohen, I. (2007). On the stationarity of Markov-Switching GARCH processes, *Econometric Theory*, 23, 485-500.
- [2] Alemohammad, N. Rezakhah, S. Alizadeh, S. H. (2015). Markov switching component GARCH model: stability and forecasting, *Communications in Statistics- Theory and Methods*, in Press. <http://www.tandfonline.com/doi/abs/10.1080/03610926.2013.841934>
- [3] Amado, C. Terasvirta, T. (2011). Conditional correlation models of autoregressive conditional heteroskedasticity with nonstationary GARCH equations, *NIPE working paper*, University of Minho.
- [4] Ardia, D. (2009). Bayesian estimation of a Markov switching threshold asymmetric GARCH model with Student-t innovations, *The in Econometric journal*, 12 (2), 105-126.
- [5] Bauwens, L. Lubrano, M. (1998). Bayesian inference on GARCH models using the Gibbs sampler, *Econometrics journal*, 1, 23-46.
- [6] Bauwens, L. Storti, G. (2009). A component GARCH model with time varying weights, *Studies in Nonlinear Dynamics and Econometrics*, 13 (2), Article 3.
- [7] Bauwens, L. Preminger, A. Rombouts, V.K. (2010). Theory and inference for Markov switching GARCH model, *Econometrics journal*, 13, 218-244.
- [8] Black, F. (1976). studies in stock price volatility changes, *Proceedings of the American Statistical Association, Business and Economics Statistics*, pp. 177-181.
- [9] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.
- [10] Chib, S. (1996). Calculating posterior distributions and model estimates in Markov mixture models, *Journal of Econometrics*, 75, 79-97.
- [11] Christofferssen, P. (1998). Evaluating interval forecasting, *International Economic Review*, 39, 841-862.
- [12] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica*, 50, 987-1007.
- [13] Engle, R. F. (1990). Discussion: stock volatility and the crash of '87, *Review of Financial Studies*, 3, 103-106.

- [14] Gelman, A. Hwang, J. Vehtari, A. (2014). Understanding predictive information criteria for Bayesian, *Statistics and Computing*, 24, 997-1016.
- [15] Glosten, L.R. Jagannathan, R. Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 48, 1779-1801.
- [16] Gonzalez-Rivera, G. (1998). Smooth transition GARCH models, *Studies in Nonlinear Dynamics and Econometrics*, 3, 61-78.
- [17] Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics*, 42, 27-62.
- [18] Grimmett, G. Stirzaker, D. (2001). *Probability and random processes*, Oxford University press, New York.
- [19] Haas, M. Mittnik, J. Paoletta, M.S. (2004). A new approach to markov-switching GARCH models, *Journal of Financial Econometrics*, 2, 493-530.
- [20] Haas, M. Krause, J. Paoletta, M.S. Steude, S.C. (2013). Time varying mixture GARCH models and asymmetric volatility, *The North American Journal of Economics and Finance*, 26, 602-623.
- [21] Hamilton, J. D. Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics*, 64, 307-333.
- [22] Kaufman, S. Fruhwirth-Schnatter, S. (2002). Bayesian analysis of switching ARCH models, *Journal of Time Series Analysis*, 23, 425-458.
- [23] Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH, *Empirical Economics* 27, 363-394.
- [24] Lancaster, P. Tismenetsky, M. (1985). *The theory of matrices*, 2nd ed, Academic press.
- [25] lubrano, M. (2001). Smooth transition GARCH models: a Bayesian approach mixture models, *Recherches Economiques de Louvain*, 67, 257-287.
- [26] Medeiros, M.C. Veiga, A. (2009). Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1), *Journal of Econometric Theory*, 25, 117-161.
- [27] Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: a new approach, *Journal of Econometrica*, 59, 347-370.

- [28] Ritter, C. Tanner, M. A. (1992). Facilitating the Gibbs sampler: The Gibbs Stopper and the Griddy-Gibbs Sampler, *Journal of the American Statistical Association*, 87, 861-868.
- [29] Zakoian, J. M. (1994). Threshold heteroskedastic models, *Journal of Economic Dynamic and Control*, 27, 577-597.

